

# USING THE OPTIMAL RANK VALUES CALCULATOR

Mike Gansecki, EPA Region 8

March 10, 2009

## Summary

The Optimal Rank Values Calculator (2006 Version) is used to identify optimal rankings for 1:1 through 1:4 non-parametric future value prediction limits, providing numerical power estimates and ratings as well. It was designed as a stand-alone Visual Basic program which can be run on most computers and fills the need for identifying maximal rank values less than the 1<sup>st</sup> or 2<sup>nd</sup> highest background values used as a prediction limit in these tests. 1:3 and 1:4 tests in particular may perform optimally (in the sense of balancing the design false positive rate and power) at lower maximal rankings under certain input conditions. In addition to providing better power, using a lower maximal value also helps to avoid problems with background data set outliers by using rank values more typical of a distribution.

The calculator operates when the user provides four inputs: 1) the design cumulative false positive error rate **DCFP**, 2) the background sample size **n**, 3) the number of tests against this background **r**, and 4) the lowest allowable ranking **j**. Please note that the calculator is designed using inverse rankings of maximum values (i.e., the absolute maximum rank of any sample set of size **n** is **j** = 1, the second highest is **j** = 2, etc.). The lowest allowable rank is set as a default to **j** = **n** (the lowest possible ranking) unless changed. Program outputs are the achievable cumulative false positive rate  $\alpha_{cum}$  at the optimal rankings, Unified Guidance power ratings (Good, Acceptable, and Low) and fractional power estimates at  $2\sigma$ ,  $3\sigma$  and  $4\sigma$  for each test. The output reports the optimal rank status for a test as "DCFP OK @ max = **j**" if the **j**<sup>th</sup> rank allows for an achievable  $\alpha_{cum}$  less than the DCFP. If the DCFP is exceeded at the maximum rank (**j** = 1), the calculator reports this as "DCFP exceeded at max=1". Additionally, the calculator provides an estimate of the total expected number of samples for **r** tests (including repeats) given a background level assumption for the test outcomes.

Two modes of operation are possible: identification of the optimal rankings and power for the inputs just described, and an identification of the achievable cumulative false positive error rates and power for the four tests when a fixed maximal rank value is specified. The following examples illustrate it's use in these two modes.

**Example 1.** Find the optimal rankings, achievable cumulative false positive error rates, and power ratings for the four tests when the design cumulative false positive error rate is **DCFP** = .002, the background sample size is **n** = 56, and the number of tests is **r** = 12

Enter the following inputs: **DCFP** = .002; **n** = 56; **r** = 12; and leave **j** = 56 as the default minimum ranking. Hit the Calculate button or Enter key. The following information is provided:

| Test Type | # Samps | Optimal Rank        | Status | Achievable CFP | Power Rating | @ $2\sigma$ | $3\sigma$ | $4\sigma$ |
|-----------|---------|---------------------|--------|----------------|--------------|-------------|-----------|-----------|
| 1:1       | 12      | DCFP exc. @ max = 1 |        | .17647         | Good         | .364        | .732      | .943      |
| 1:2       | 12      | DCFP exc. @ max = 1 |        | .00713         | Good         | .142        | .544      | .892      |
| 1:3       | 12      | DCFP OK @ max = 2   |        | .00147         | Good         | .162        | .628      | .938      |
| 1:4       | 13      | DCFP OK @ max = 5   |        | .00172         | Good         | .286        | .791      | .979      |

## Summary

These results indicate that both the 1:1 and 1:2 tests would exceed the target cumulative false positive error rate even at the maximum ranking  $j = 1$ , the 1:1 test by a substantial margin. The 1:3 test is optimized at the 2<sup>nd</sup> highest maximum  $j = 2$ , while the 1:4 test can use the 5<sup>th</sup> highest value at optimality. For the latter two tests, superior power is shown with the 1:3 and 1:4 tests at the three  $\sigma$  levels, although all four tests have "Good" power. The number of expected samples and repeats is 12 for all but the 1:4 test (13). Only a modest increase in samples would be expected.

**Example 2.** Identify the achievable false positive error rates and power at the maximum rank value, using the same  $n$  and  $r$  inputs as **Example 1**.

In order to find the achievable false positive error rates using the maximum value, replace the design input error rate **DCFP** with 1.0, and the allowable ranking  $j$  with 1. The outputs generated are:

| Test Type | # Samps | Optimal Rank Status | Achievable CFP | Power Rating | @ 2 $\sigma$ | 3 $\sigma$ | 4 $\sigma$ |
|-----------|---------|---------------------|----------------|--------------|--------------|------------|------------|
| 1:1       | 12      | DCFP OK @ max = 1   | .17647         | Good         | .364         | .732       | .943       |
| 1:2       | 12      | DCFP OK @ max = 1   | .00713         | Good         | .142         | .544       | .892       |
| 1:3       | 12      | DCFP OK @ max = 1   | .00037         | Acceptable   | .059         | .409       | .843       |
| 1:4       | 12      | DCFP OK @ max = 1   | .00002         | Low          | .026         | .310       | .798       |

It can be seen that at the maximum ranking (or any other fixed maximum rank), the lowest achievable CFP occurs with the 1:4 test and increases as the  $m$  of the  $1:m$  test decreases. Power is inversely related and is greatest with the 1:1 test and decreases with increasing  $m$ , also as shown by the change in ratings. These outcomes do reflect the superiority of optimizing higher  $1:m$  tests.

Using these two calculator modes, the user can experiment with various inputs to find the best overall test design from among these non-parametric prediction limit test choices. The narrative below provides greater detail on defining the appropriate inputs, interpreting outputs, and how the calculator functions.

## Introduction

The Optimal Rank Values Calculator (2006 Version) is a stand-alone software program used for the design and evaluation of certain non-parametric prediction limit statistical tests. Written in Visual Basic 6, the program file "OptRank06.exe" has a total size of about 460K and should run on most Windows operating systems. An initial version of this program was provided with the 2004 draft of the Unified Guidance, and has since been updated. The attachment to this summary provides greater detail on calculator algorithms and performance.

The program was written and developed by Mike Gansecki, EPA Region 8. Steve Burkett, also of EPA Region 8, provided assistance in its formulation. For any further questions on the calculator, please contact Mike at [gansecki.mike@epa.gov](mailto:gansecki.mike@epa.gov) or at: 303-312-6150.

## **Purpose of the Calculator**

The Optimal Rank Values Calculator provides design information for certain 1:*m* non-parametric future value prediction limits (PL) tests. The calculator identifies the optimal background maximal rank value for 1:1 to 1:4 PL tests which can meet a design cumulative false positive error for a specific background sample size, the number of tests against that background, and any limitations on the available maximal rank as inputs. It also provides fractional power estimates at 2, 3, and 4  $\sigma$  levels above a standard normal background for each 1:*m* test, and rates power consistent with the Unified Guidance methodology.

The calculator is intended to supplement information in the EPA RCRA document, “Statistical Analysis of Ground Water Monitoring at RCRA Facilities– Unified Guidance”, currently under development as of 2009. The Unified Guidance document covers six different non-parametric prediction limit tests recommended for detection monitoring– 1:2, 1:3, 1:4, and a modified California plan future values tests, as well as 1:1 and 1:2 size 3 median tests. The calculator evaluates only the 1:2 through 1:4 future values guidance tests (the 1:1 test is included since it follows a common algorithm and can serve for comparison purposes). Particularly with the higher 1:*m* tests, the optimal balance between meeting both a design cumulative false positive error rate and sufficient power may only be achieved with lower background prediction limit maximal ranks. The modified California and median non-parametric guidance tests have statistical characteristics which limit the likelihood of optimal maxima to the largest or 2<sup>nd</sup> largest rank value meeting both false positive/power criteria, and are sufficiently covered in existing guidance tables.

## **Statistical Testing**

RCRA groundwater monitoring regulations require statistical comparison of future downgradient (compliance) well samples against background for a variety of constituents, to determine if the downgradient monitoring well(s) indicate a release. Generally, a statistically significant increase above background is indicative of such a release. The background data for a given constituent may be obtained from upgradient wells, pooled or individual historical compliance wells, depending on statistical characteristics as described in the Unified Guidance.

Parametric and non-parametric prediction limits are among the statistical test options which can be used for detection monitoring. A prediction limit identifies some statistic from a background data set used as the basis of comparison for future samples or sample statistics from compliance monitoring well data. Given that most comparison tests are intended to evaluate increases, a one-way upper prediction limit is most frequently used. For intrawell testing, this can be a two-sample comparison, but is often extended as multiple comparison tests if data from a number of compliance monitoring wells at different time intervals are tested against the same background prediction limit.

The value of the prediction limit is determined by a number of factors including the test design, the background sample size, the number of tests against this background limit, and the level of test confidence desired. The prediction limit is derived solely from background data, and it is initially assumed under the null hypothesis that compliance well data follow the same pattern. Prediction limit test outcomes are based on how a given number of future well compliance samples are expected to fall within this limit under background conditions. Certain future sample data patterns are considered statistically significant indications of a release or exceedance, while others falling within the upper prediction limit are considered “in-bounds” (i.e., not significantly different from background)..

When conditions warrant, it may be necessary to apply non-parametric testing methods. Among the most powerful are 1: $m$  prediction limit tests (the term 1: $m$  or 1-of- $m$  refers to the formal use of repeat testing from 1 to  $m$  resamples at a monitoring well testing location). In this type of test, future downgradient samples (and repeats as necessary) are compared to some fixed maximal value of the background data set used as a prediction limit. For a 1:2 prediction limit test, both the initial sample and resample must exceed the prediction limit to be declared an exceedance. For a 1:4 test, 0, 1, 2 or 3 exceedances out of 4 are still considered “in-bounds”; only 4:4 exceedances are considered significantly different from background. While this may seem excessively lenient from a regulatory perspective, the calculator is designed to provide optimal comparison conditions for higher order 1: $m$  tests (i.e., lower maxima) which meet the same overall cumulative false positive objective. As will be seen, power actually increases for higher level 1: $m$  tests under the same input conditions when maximal values are optimized.

The virtue of prediction limits is that they are highly flexible and can generally be adjusted to either meet or approximate exact confidence levels for a wide range of facility-specific conditions. While parametric prediction limits using  $\kappa$ -multiples can obtain exact confidence levels, non-parametric prediction limits are based on fixed achievable false positive error rates for a given test, the maximal value used as a prediction limit, background sample size and other design factors. Particularly with the use of this Optimal Rank Values Calculator, it is possible to approximately meet the SWFPR objectives with non-parametric prediction limits by careful consideration of various design inputs including the level of the 1: $m$  test.

The Unified Guidance uses a statistical design process for detection monitoring at a facility which incorporates the pre-determined target annual site-wide false positive rate (**SWFPR**) and identifies minimum satisfactory power criteria (the ability to discriminate certain increases above background for a single monitoring well-constituent test). While the details of this design process are too involved to be presented here, the following is a brief summary of how calculator inputs would be defined by the design factors.

To make effective use of prediction limit testing in detection monitoring site design, a target individual test false positive rate and cumulative false positive rate for comparisons to a single background must be established. These in turn, depend on the pre-selected overall facility-wide false positive error rate (**SWFPR**) for some relevant time period. The Unified Guidance

recommends that a facility-wide false positive rate be applied to the total number of tests in a given calendar year, and has suggested an **SWFPR** = .1 or 10%. Two approaches can be used, based on either an exact Binomial distribution formula for the probability of at least one or more occurrences given a single test error rate or the Bonferroni approximation to the first method. The Binomial approach is preferred, since it follows the approach in the Unified Guidance (March 2009). The applications here are for non-parametric prediction limit tests, considering either future values tested against a common constituent background data set (interwell) or tested against historical data from the same well (intrawell).

The *total* number  $g$  of sitewide annual independent statistical tests which contribute to the overall SWFPR must first be identified.  $g$  is the product of the number of compliance wells [ $w$ ], the number of valid monitoring constituents [ $c$ ], and the number of required evaluations per year [ $n_E$ ]:

$$g = w \times c \times n_E$$

The individual design false positive error rate for a single test ( $\alpha_{test}$ ) is then:

$$\alpha_{test} = 1 - (1 - \text{SWFPR})^{1/g} \quad \text{Equation 1.}$$

Once the single test rate  $\alpha_{test}$  is known, the cumulative false positive  $\alpha_{cum}$  for any of  $r$  tests can be calculated by reversing the equation:

$$\alpha_{cum} = 1 - (1 - \alpha_{test})^r \quad \text{Equation 2.}$$

Using a combination of these two equations will allow for the calculation of the cumulative false positive error rate for the two most important testing comparisons— interwell and intrawell testing.

Because of the nature of the mathematical theory behind non-parametric prediction limits, tests are evaluated on a single background constituent sample size, considering the number of tests against this background. To properly utilize the calculator, the target cumulative false positive error rate must first be established for either interwell or intrawell comparisons.

In interwell testing, it is most common to evaluate all compliance well data tests against a single common background. In this event, the effective number of tests is then  $w$  the number of compliance wells times  $n_E$  the number of evaluations per year or  $r = w \times n_E$ . The cumulative false positive rate  $\alpha_{cum}$  for this number of tests given the single test false positive rate can be calculated using Equation 2. An alternate way of calculating the cumulative interwell error rate is to apportion the full **SWFPR** =  $\alpha = .1$  equally to the  $c$  constituents:

$$\alpha_{cum} = 1 - (1 - \text{SWFPR})^{1/c} \quad \text{Equation 3.}$$

The results using either Equations 2 or 3 should be identical.

For intrawell comparisons, future data for each compliance well are tested against the single historical background sample set from that well. In that case, there are only  $r = n_E$  tests. Equation 2 can be used with the single test false positive error rate  $\alpha_{test}$  to obtain the target cumulative error:

$$\alpha_{cum} = 1 - (1 - \alpha_{test})^{r = n_E}$$

The Optimal Rank Values Calculator software uses  $n$ ,  $r$  and  $\alpha_{cum}$  as inputs to identify optimal non-parametric rank values (provided as inverse rankings, i.e., the maximum rank of a background sample = 1) which can attain single constituent cumulative error rates ( $\alpha_{cum}$ ) for 1:1 through 1:4 tests of future values. The calculator interface identifies the design cumulative false positive error rate for a single background prediction limit as **DCFP**, instead of  $\alpha_{cum}$ . Power is independently estimated once the optimal ranks have been determined.

It is also possible to utilize the Bonferroni approximation to the Binomial formulas, since the error rates are nearly linear. The **SWFPR** can be divided by  $g = w \times c \times n_E$  or  $.1/g$  to obtain the single test false positive error rate. A cumulative error rate  $\alpha_{cum}$  can be obtained by multiplying the single test error rate by  $r$  or:  $\alpha_{cum} = r \times \alpha_{test}$ . Since the Binomial approach is exact, it is preferred for calculator use. However, the differences using the Bonferroni approximation are generally very minor.

## CALCULATOR OPERATION

Figure 1 provides an reasonable approximation of what is seen in the calculator window screen. There are three blocks or groups: Input values, Status & Power outputs, and Additional Information outputs. Five function keys are provided– Calculate, Print, Reset, Exit and “About this calculator.”

### Input Values

Four related inputs must be identified for the calculator to function. Each input box is described below:

**Design Cumulative False Positive Error (DCFP)**– As described above, this rate is the design cumulative error false positive error ( $\alpha_{cum}$ ) for  $r$  tests against a background of size  $n$ . The program only accepts fractional values from  $> 0$  to 1.0, and an error message will result if other values are entered. For example, if 8 different constituents with different size background databases will be tested interwell, the single-constituent error rate is the overall .1 cumulative error partitioned to the  $c = 8$  constituents (.01308) or divided by 8 = .0125 using the Bonferroni approximation. This error is cumulative, since

it must account for the number of individual comparison tests to this background, described below.

**Background sample size ( $N$ )** – This is the background sample size defined for the  $r$  tests at the **DCFP** rate, and can represent pooled up- and downgradient well or intrawell data, depending on their statistical characteristics. The program only accepts positive integers of 2 or greater. Input values as high as 1000 have been used without causing program failure; however the expected range is from 4 to 300. If a site can aggregate multiple well upgradient or other data, larger background sample sizes are possible. Typical single-well background sample sizes are more like 4 to 25.

**Number of comparison (tests) to the same background ( $r$ )**–. The number of comparison tests is against a single background prediction limit in a specified time period. Repeat test samples are not included in this value, and are accounted for in the program as individual 1: $m$  results. The program accepts integer values of 1 or greater. Values as high as 1000 have been used without causing the program to terminate; however, typical ranges are from 1 to 500. If sample constituent data from 20 downgradient monitoring wells will be compared to this single background prediction limit twice a year, the number of annual comparison (tests) is  $r = 20 \times 2 = 40$ . If each downgradient well set were compared to its own historical background (intrawell comparisons), the number of comparisons would then be  $r = 2$  (the cumulative false positive **DCFP** and background sample size  $n$  would also need to be adjusted).

**Minimum allowable rank value ( $j$ )**– This is the lowest allowable maximal value of the background data set which can be used for comparison. The calculator uses inverse ranks, i.e., the maximum value  $j=1$ , the 2<sup>nd</sup> highest value  $j=2$ , etc. The program sets the lowest rank by default to the smallest background value ( $n$ ), but can be adjusted to some other integer value from 1 to  $n-1$ . This could occur if the non-parametric background data contain non-detect values. If a background sample size of 20 had 75% non-detects,  $j=5$  would be entered as the lowest maximum allowable inverse rank value

The calculator can be operated in other than an optimizing mode to provide exact cumulative false positive error rates for a fixed maximal value of a background sample set for the four 1: $m$  tests. This is done by setting the DCFP to 1.0 and identifying the desired minimum allowable inverse rank value  $j$ .

While the calculator is designed to handle typical inputs even from a very large RCRA facility, there are upper performance limits. A run-time error will occur for certain combinations of large **DCFP**,  $n$  and  $r$  values. The Attachment provides a graph identifying levels for the three inputs which can result in a run-time error. Calculator operation times are typically only a few seconds, but can be longer if  $n$  and  $r$  are large (e.g.  $> 500$ ).

## Function Keys

Calculate

Clicking this button initiates the program when all four valid inputs are entered. The Return key will also perform this function as well as hitting [alt] and C keys. The program calculates successive maximal values (beginning with the absolute maximum  $j=1$ ) for a given sample size and number of comparisons up to the background sample size or specified maximum limit until the optimal cumulative false positive for a given  $1:m$  test is reached. The optimum is the lowest maximal value which still meets the DCFP cumulative false positive input. If the 1<sup>st</sup> maximal value exceeds the DCFP for a test type, this is reported.

In program calculations, each of the four  $1:m$  tests is evaluated in turn. After all  $1:m$  test optimal maxima are identified, the program then calculates relative power based on a standard normal distribution for background and normal future samples at 2, 3 & 4 sigma levels). Finally, the program identifies a power rating of "Low", "Acceptable" or "Good" power depending on whether none, one or both of the EPA Reference power criteria are met. All output value boxes are filled.

Print

Clicking this button or [alt]- P keys sends a copy of the overall Windows screen output to a local or line printer.

Reset

Clicking this button or [alt]- R keys will clear all of the input and output boxes for new inputs and calculation. Alternatively, if it desired to change only a single input, simply re-enter the desired input value and click on the Calculate button or Enter key.

Exit

Clicking this button or [alt]- x keys closes the Visual Basic program.

About this calculator

Clicking this button or [alt]- A keys provides a summary of main calculator features, input information and program development.

## Outputs– Status & Power

For a set of input conditions, the information under "DCFP Status @ Opt. Maximum" identifies whether or not the design cumulative false positive was exceeded or is satisfactory



(OK). It identifies the inverse rank maximal value for each 1:*m* test which is optimal in meeting the cumulative false positive (unless exceeded at the maximum *j* = 1).

The following example illustrates the use of the calculator in optimizing mode. Assume **c** = 10 monitoring constituents; **w** = 10 compliance wells at a facility to be tested semi-annually (**n<sub>E</sub>** = 2), with 65 pooled background samples available for interwell testing; and an overall **SWFPR** = .1. Then **r** = 10 × 2 = 20 and **DCFP** = 1-(1-SWFPR)<sup>1/c</sup> = .0105. Further assume that all background data are above detection.

In Figure 1, the following input conditions are shown: the DCFP cumulative false positive error = .0105; background sample size = 65; number of comparisons = 20; with no constraints on the lowest rank (default minimal allowable rank value = 65). For the 1:1 test, the cumulative false positive status is "DCFP exceeded at max = 1". For the 1:2 test, the cumulative false positive status is "DCFP OK at max = 1". For the 1:3 test, the 4<sup>th</sup> highest value is optimal, while the 8<sup>th</sup> highest value is optimal for the 1:4 test.

Unified Guidance power ratings for each of the tests are provided in the last column. Although rated “Good”, the 1:1 test power results are moot since this test significantly exceeds the design DCFP (an achievable cumulative false positive of .235 using the maximum rank). The rating is “Acceptable” for the 1:2 test at the 1<sup>st</sup> maximum, since the 4σ EPA Reference Criterion shown at the bottom of the calculator was equaled or exceeded. The power ratings for both the 1:3 and 1:4 tests at their respective optimal maximal ranks are "Good", since both the 3 & 4 σ criteria were exceeded. These category ratings follow that of the 2009 draft Unified Guidance. The reference rating values shown at the bottom of the calculator outputs are based on an approximation methodology for non-central t-distribution normal power evaluations used in the calculator. A more detailed explanation of this methodology is found in the Attachment.

### Additional Information

The first column after the 1:*m* test listings is identified as the “# of Samps”. These represent the expected number of total comparison samples including repeats under background conditions for the four 1:*m* test types. Since the input value for the number of single comparison tests was 20, this would be expected number as well for the 1:1 test. The same (rounded) number of expected samples is shown for the 1:2 test. For the 1:3 test, 21 samples (including repeats) would be expected and 23 for the 1:4 test type. The estimates are based on the individual P values in the next column. For the 1:4 test, the 8<sup>th</sup> maximum value out of 65 would have a single comparison confidence level of .8846 or a 1 - .8846 = .1154 probability of being exceeded by future samples from background. The individual P value is calculated as  $IndP = (n-j+.5)/n$ , where **n** is the background sample size and **j** is the optimal maximal rank. The expected numbers of samples (rounded to the nearest integer) for **r** comparisons and 1:*m* tests are calculated as:

$$\#Tests = r \cdot \left( \sum_{i=0}^{m-1} (1 - IndP)^i \right)$$

Achievable cumulative false positive error rates given the input conditions and optimal maximal rank value for each 1:*m* test type are found in the next column of Figure 1. The software program uses double-precision numbers in calculating these output values to five decimal places.

The final three columns provide the estimated fractional power at 2, 3, and 4  $\sigma$  mean levels above background for each of the four 1:*m* tests, based on normal distribution assumptions. As can be noted, the 1:4 test is the most powerful under optimal rank conditions. Power values are generated from an approximation program based on Davis & McNichols' basic equation for parametric normal prediction limit tests. More complete explanations for both the algorithm used to calculate the 1:*m* cumulative false positives and the power approximations are found in the attachment to this summary.

### **Benefits of the Calculator**

The heart of the calculator program is the algorithm for calculating the optimal maximal cumulative false positives. In the 2009 Unified Guidance, a substantial number of appendix tables (**Appendix D, Tables 19-19 to 19-24**) are needed to cover only the absolute and second highest maximum values used in these non-parametric tests. It is impractical to generate even more tables for lower maxima; yet especially for higher level 1:*m* tests, the optimal values calculated with this program suggest much lower maxima under certain conditions. As the above example demonstrates, the 8<sup>th</sup> highest maximal value was optimal for the 1:4 test. Particularly where background sample sizes are very large (e.g.,  $n > 100$ ), one might expect the 5<sup>th</sup>, 10<sup>th</sup>, 17<sup>th</sup> or even the 20<sup>th</sup> highest maximum to be the optimal maximum for the higher level 1:*m* tests under certain input conditions.

If one wished to generate cumulative false positive error and power information for a given background sample size and the number of comparison tests at a fixed maximal value (e.g., the highest background value), the calculator allows for this option. By specifying the cumulative false positive input **DCFP** = 1 and the minimum detected value *j*, and varying the other two variables, the calculator outputs will provide the achievable cumulative false positive error and power for each of the four tests at that maximal rank. Using the example background sample size and number of comparison tests from Figure 1 with the **DCFP** set to 1 and the minimal rank value *j* = 1, the results are shown in Figure 2. At the maximum value used as a prediction limit, the 1:4 test demonstrates the lowest achievable cumulative false positive error, but also the lowest power. This example illustrates the value in optimizing the maximal rank value used as a prediction limit. Higher order 1:*m* tests are less powerful at a given fixed rank value, but more powerful tests when optimized.

The cumulative false positive error values based on the calculator algorithm have been checked against existing tables in statistical texts<sup>1</sup> and the guidance, and are accurate to four or five significant places. While the fractional power estimates are approximate, they are accurate enough to closely match the ratings provided in the guidance. By providing the actual power fractional estimates, more latitude is available should these reference criteria be modified. Fractional power levels also allow for determining effect-size power. The attachment describes a comparison of power ratings and estimates using the calculator with the 2004 draft Unified Guidance tables and Monte Carlo data, and one approach to calculating specific effect size power.

Using this optimizing calculator with typical input variable conditions, it can be more easily seen that higher level 1: $m$  tests provide greater relative power with no loss of false positive control. Regulatory agencies may find it easier to allow use of these higher level type tests. The calculator also offers an easy way to evaluate 1: $m$  non-parametric prediction limit tests in overall detection monitoring system design.

Figures and Attachment

---

<sup>1</sup> For example, Gibbons, Robert D., Statistical Methods for Groundwater Monitoring, John Wiley & Sons, Inc., 1994, Tables 2-5 through 2-12, pp.52-67. Since the values are given as confidence probabilities of lying below a maximum value, simply subtract 1- cumulative false positive of the Calculator values.

**Figure 1. Window Screen for Optimal Rank Calculator Program (Optimization Mode Example)**

# Optimal Rank Values Calculator

Input Values Here:

|  |       |
|--|-------|
| Design Cum. False Positive Error [DCFP]: | .0105 |
| Background sample size [n]:              | 65    |
| Number of comparisons to same bkg.[r]:   | 20    |
| Minimum allowable rank value [j] *:      | 65    |

\* The default value is the lowest allowable inverse rank from the Background Sample Size above. You may overwrite this default with another rank value.

Calculate

Print

Reset

Exit

About this calculator

## Outputs

### Status & Power

| Test | DCFP Status @ Opt. Maximum | Power Rating |
|------|----------------------------|--------------|
| 1:1  | DCFP exceeded at max = 1   | Good         |
| 1:2  | DCFP OK at maximum = 1     | Acceptable   |
| 1:3  | DCFP OK at maximum = 4     | Good         |
| 1:4  | DCFP OK at maximum = 8     | Good         |

### Additional Information

| Test | # of Samp | Ind P | Achievable CFP Error | Power @ |       |       |
|------|-----------|-------|----------------------|---------|-------|-------|
|      |           |       |                      | 2 Sig   | 3 Sig | 4 Sig |
| 1:1  | 20        | .9923 | .23529               | .344    | .716  | .938  |
| 1:2  | 20        | .9923 | .00883               | .126    | .519  | .881  |
| 1:3  | 21        | .9462 | .00787               | .286    | .769  | .973  |
| 1:4  | 23        | .8846 | .00755               | .392    | .860  | .989  |

EPA Reference Criteria:  $\geq .539$   $\geq .814$

**Figure 2. Window Screen for Optimal Rank Calculator Program (Fixed Rank Mode Example)**

# Optimal Rank Values Calculator

Input Values Here:

|  |    |
|--|----|
| Design Cum. False Positive Error [DCFP]: | 1  |
| Background sample size [n]:              | 65 |
| Number of comparisons to same bkg.[r]:   | 20 |
| Minimum allowable rank value [j] *:      | 1  |

\* The default value is the lowest allowable inverse rank from the Background Sample Size above. You may overwrite this default with another rank value.

Calculate

Print

Reset

Exit

About this calculator

## Outputs

### Status & Power

| Test | DCFP Status @ Opt. Maximum | Power Rating |
|------|----------------------------|--------------|
| 1:1  | DCFP OK at maximum =1      | Good         |
| 1:2  | DCFP OK at maximum = 1     | Acceptable   |
| 1:3  | DCFP OK at maximum = 1     | Acceptable   |
| 1:4  | DCFP OK at maximum = 1     | Low          |

### Additional Information

| Test | # of Samp | Ind P | Achievable CFP Error | Power @ |       |       |
|------|-----------|-------|----------------------|---------|-------|-------|
|      |           |       |                      | 2 Sig   | 3 Sig | 4 Sig |
| 1:1  | 20        | .9923 | .23529               | .344    | .716  | .938  |
| 1:2  | 20        | .9923 | .00883               | .126    | .519  | .881  |
| 1:3  | 21        | .9923 | .00040               | .049    | .381  | .829  |
| 1:4  | 23        | .9923 | .00002               | .020    | .283  | .780  |

EPA Reference Criteria: >= .539 >= .814

# ATTACHMENT

## Development of 'OPTIMAL RANK VALUES CALCULATOR' Program for 1-of- $m$ Non-Parametric Prediction Limit Tests— 2006 Version May 18, 2007

### Introduction

In developing this software program to balance false positives and negatives utilizing 1-of- $m$  non-parametric tests, it was necessary to evaluate the cumulative false positive level for various combinations of input 1: $m$  test type, background sample size, the number of comparison tests, and the maximum value used for comparison. The cumulative false positive is evaluated for individual constituents (against a single background database). The initial formula used was a variant of the multivariate hypergeometric distribution described by Hall, Prarie & Motlagh (1975) and Chou & Owen (1986) provided in Gibbon's 1994 book<sup>2</sup>. Using slightly different notation as  $m$  the number comparisons within the 1: $m$  test type,  $n$  sample size, and  $r$  the number of comparison tests, the cumulative confidence level,  $\gamma$  is given as follows:

$$\gamma = \frac{n}{mr + n} \sum_{j_1=1}^m \sum_{j_2=1}^m \dots \sum_{j_r=1}^m \frac{\binom{m}{j_1} \binom{m}{j_2} \dots \binom{m}{j_r}}{\binom{mr + n - 1}{\sum_{i=1}^r j_i + n - 1}} \quad \text{Equation (2.2)}$$

where  $\binom{m}{j}$  refers to the number of combinations of  $m$  things taken  $j$  at a time. As presented, the formula evaluated the maximum sample value only. In subsequent discussions, Gibbons indicated that the number of terms increased roughly as a factor of  $m^r$ , making exact calculations of larger numbers of comparison tests impractical. But in working with small  $n$  and  $r$  applications, it was found that only  $(m-1)r+1$  distinct terms were needed within a single summation for a solution, a much more tractable problem. Further investigations showed that  $\gamma$  could be calculated as a single summation utilizing four major terms (A-D) for any maximum value  $MX$  (where the maximum value =1 and lesser maxima in inverse order):

$$\gamma = D(m, n, r, MX) \sum_{i=0}^{(m-1)r} \frac{A(m, r)_i \times B(MX)_i}{C(m, n, r, MX)_i}$$

with  $D$  a constant for any given combination of  $m$ ,  $N$ ,  $r$ , and a maximum value  $MX$ . Each of the four terms is defined below:

<sup>2</sup>

Gibbons, Robert D., Statistical Methods for Groundwater Monitoring, John Wiley & Sons, 1994, p.37, Equation 2.2

### D Term:

$$D = \frac{(n)!}{(n-MX)!} \times \frac{(n+mr-MX)!}{(n+mr)!}$$

As an example, for  $MX=2$  and  $m=3$ ,  $D = \frac{n \times (n-1)}{(n+3r) \times (n+3r-1)}$ .

### B Term

B is an expansion of the series  $(MX+i)!$ , as  $i$  varies from 0 to  $(m-1)r$ . It is defined as:

$$B_i = \frac{(MX-1+i)!}{(MX-1)!}$$

If  $MX=2$ , then the series of terms for  $i = 0$  to 4 is: 1, 2, 6, 24..... etc.

### C Term

In Equation 2.2 above, the upper term  $mr+n-1$  in the combination  $\left( \sum_{i=1}^r j_i + n - 1 \right)$

corresponds more generally to the denominator C defined as follows: if  $L = mr+n-MX$ , then:

$$C_i = \frac{(mr+n-MX)!}{(mr+n-MX-i)!} \text{ or } \frac{L!}{(L-i)!}$$

For  $k=3$ ,  $n=5$ ,  $m=2$ , and  $MX=1$ , then if  $L = 10$ , for the first four terms  $i = 0$  to 4,  $C_i = 1, 10, 10 \cdot 9, 10 \cdot 9 \cdot 8$ , or 1, 10, 90, 720, ...etc.

### A Term

The A terms are coefficients and the most difficult to mathematically formulate. They are variants of a binomial/polynomial expansion and the  $(m-1)r+1$  terms are defined by the equation  $[(1+a)^m - a^m]^r$  as follows:

| Test Type | Expansion            | Number of Terms |
|-----------|----------------------|-----------------|
| 1:1       | $1^r$                | 1               |
| 1:2       | $(1+2a)^r$           | $r+1$           |
| 1:3       | $(1+3a+3a^2)^r$      | $2r+1$          |
| 1:4       | $(1+4a+6a^2+4a^3)^r$ | $3r+1$          |

Setting  $a=1$  results in the exact coefficients for each of the  $(m-1)r+1$  terms. As an example of a 1:2 test with  $r=5$  comparison tests, the six coefficients from the  $(1+2a)^5$  expansion would be: 1, 10, 40, 80, 80, & 32. Computationally, the coefficients are best generated using recursive programming techniques.

The overall formula for generating  $\gamma$  is described in the following equation. For the 1:1 test case, only the D coefficient is calculated, since there is only a single summation term = 1:

$$\gamma = \frac{n! \times (mr + n - MX)!}{(n - MX)! (mr + n)!} \sum_{i=0}^{(m-1)r} \frac{(MX - 1 + i)! (mr + n - MX - i)!}{(MX - 1)! (mr + n - MX)!} [a(m, r)]_i$$

where  $a(m, r)_i$  are the coefficients as described above.

### Programming Considerations

Certain software design features were felt important for program applications in calculating the cumulative false positive for the various 1-of- $m$  tests. Since the number of comparison tests  $r$  is to a single background sample set, the program should be able to calculate using inputs of up to 200-500 comparison tests and background sample sizes as reasonable maxima.

Secondly, the program should perform fairly quickly, particularly for the 1:4 test involving the greatest number of computational terms. The program should be able to assess maximal values lower than the simple sample maximum or 2<sup>nd</sup> largest, particularly for higher order 1:3 or 1:4 tests.

A final desirable feature was that the recursive program should be as stable as possible under the likely input assumptions of  $r$ ,  $MX$ , and  $n$ . A recent paper by Davis and McNichols<sup>3</sup> provided a comprehensive presentation of 1-of  $m$  and other non-parametric prediction limit tests including a simplified algorithm for calculating the cumulative false positive error. While an exceptionally quick program, the use of (+) (-) terms has its own problems of regions of instability or chaotic behavior. Their paper suggests that this algorithm works best in ranges where  $r < 3n$ . Not all maximum values could be accurately calculated even within this range.

The 'Optimal Rank Values Calculator' program was developed following the logic of the above  $\gamma$  derivation, although calculations are developed in a recursive manner for all four components. The program was initially written in DOS-BASIC, and then converted to a Visual Basic 6.0 format. In program terms, certain variable name differences occur. The number of comparisons is entered as "R" versus "r", N remains the sample size, MX the inverse maximum value rank. Instead of " $m$ " in 1: $m$  comparisons within a given test type, the program uses "Q".

---

<sup>3</sup> Davis, Charles B., and McNichols, Roger J., "Simultaneous Nonparametric Prediction Limits", Technometrics, May 1999 pp. 89 - 101.



## **‘Optimal Rank Calculator’ Program Cumulative False Positive Error and Optimization**

As shown on the flow diagram in Figure 1, the ‘Optimal Rank Values Calculator’ program evaluates each of the four 1-of-Q test situations to find the lowest achievable maximum value of the sample N data set which meets a pre-specified or design input cumulative false positive error limit. This design cumulative error rate is based on a single-constituent, single background size comparison for R tests. As shown on the diagram, inputs are background sample size (N), the number of comparison tests to a single background sample set (R), a pre-specified design cumulative false positive error (AL), and whether a lower bound maximum limit is placed on the search.

Beginning with the 1:1 test comparison in Figure 1,  $Q = 1$  and the initial maximum value is set to 1. The dark shaded box defining the “cum( $\alpha$ )” is the heart of the program and is shown in detail below in Figure 2. The program then evaluates whether the calculated cum( $\alpha$ ) is greater than the pre-specified AL or  $\alpha_c$ . If not, the second maximum is evaluated, etc. If maxima greater than 1 are found to exceed AL, the previous maximum and cum( $\alpha$ ) are used as the optimum fit for that 1:Q test. If the maximum value exceeds the cumulative false positive level, the program reports this as "DCFP exceeded at max = 1".

The program also evaluates whether any given maximal rank exceeds a pre-specified lower allowable rank value and reports the values at that rank, indicating that it was constrained by a lower limit. For increased efficiency, the initial maximal value for a 1:Q test is set at the previous optimal rank for the Q-1 test as indicated in a Figure 1 footnote, since the cumulative false positive is always lower for a given set of program inputs at a higher 1:Q test. Power (described in the next section) is evaluated for all 1:Q tests at the input N and the optimal MX for  $Q = 1$  to 4. After assessing the  $Q=4$  condition, the program prints all of the cum( $\alpha$ ), optimal maxima, and power information.

Figure 2 is a flow diagram of that portion of the calculator program which calculates the cumulative false positive error for single inputs of N, R, and MX. The C denominator term mentioned above is calculated as  $L(Q)$  for a given 1:Q test. The program works with an  $F(i, j)$  matrix of  $i$  rows and  $j$  columns. The number of rows equals  $Q$  and the number of columns varies from  $Q$  to  $RP(Q) = (Q-1)R+1$ . Next, the D term is calculated for a given  $Q$ , with a variable ASUM( $Q$ ) set to zero. If  $Q=1$ , the recursive portion of the program is skipped and  $ASUM(1)=1$ . Then  $\gamma_{1,1} = P(1) = D(1) \cdot ASUM(1)$ . The cumulative error  $ER(Q) = 1 - P(Q)$ . For 1:2 through 1:4 tests, the full  $F[Q, (Q-1)R+1]$  matrix is first set to all zeros.  $i$  is then used as a counter for  $R$  steps through the recursion process. At  $i = 1$ , initial terms are provided in the first row [ $F(1, 1)$  to  $F(1, Q)$ ] which correspond to AB/C elements described above. As an example for  $Q=3$ , initial conditions are  $F(1, 1)=1$ ,  $F(1, 2)=3MX/L$ , and  $F(1, 3)=[3MX \cdot (MX+1)]/[L \cdot (L-1)]$ .

In order to avoid extensive unnecessary multiplications against zero and increase program speed, two additional variables  $IP(Q)$  and  $IP2(Q)$  are defined at each step from  $i$  to  $R$ . Gradually, the  $Q \times IP(*)$  F-matrix expands with each successive  $i$  until it reaches  $Q \times (Q-1)R+1$  rows and

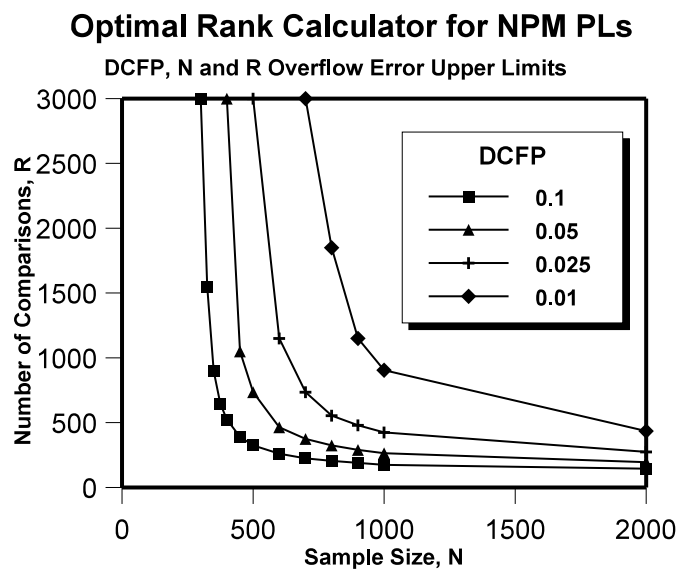
columns at the Rth iteration. Within each i iteration, a second row is defined by a multiplier against the first row contents, with additional row multipliers for each Q against the first row. The numerical multipliers correspond to the  $[(1+a)^m - a^m]^r$  coefficients in the formula described above..

Once the Qth row is calculated, the program then sums the first to the Qth row for each column. The summed values are reentered into the first row of the F(i, j) matrix and the process repeated R times. When i = R, the summed column values are then summed across the row to generate a total value as ASUM(Q). This value is then multiplied against D(Q) to obtain the cumulative confidence level P(Q), and the error ER(Q). In the full 'Optimal Rank Values Calculator' program,  $ER(Q) = \text{cum}(\alpha_c)$  for each 1:Q test type.

Figure 3 provides direct calculations for the program algorithm using an example with small input values. The figures better illustrate the recursive nature of the program. The cumulative error for 1:1 to 1:4 tests is obtained from inputs N=5, R=3, and MX=2. In the example for the 1:2 test, where D(2)=2/11, RP(2)=4, and L(2)=9, the full summary equation would be:

$$P2 = \frac{2}{11} \times [1 + 6 \times \frac{2}{9} + 12 \times \frac{2 \times 3}{9 \times 8} + 8 \times \frac{2 \times 3 \times 4}{9 \times 8 \times 7}] = \frac{2}{11} \times [1 + \frac{12}{9} + 1 + \frac{8}{21}]$$

The program design can result in fairly long running times when R and N are large, although typical applications require only a few seconds. Given the present software structure, there are upper input limits beyond which an overflow error will occur. The graph below provides limits for DCFP, N and R inputs. N & R values to the right and above a given DCFP curve will produce an overflow error (e.g., at  $\alpha = .1$ , combinations of N and R above 300-500 may produce an error). However, a realistic application of this program would utilize a smaller DCFP, with considerably higher N and R input overflow limits.



Cumulative false positive program results were checked against the eight pages of tabular values in Gibbon's text containing tables for 1:1 through 1:4 tests up to N=100 and R=100 using the maximum or 2<sup>nd</sup> largest value. The present algorithm was correct to four decimal places for all inputs checked.

### **'Optimal Rank Values Calculator' Power Evaluations**

Once the program has identified the appropriate inverse maxima for each of the 1:*m* tests, the program evaluates power based on normal distribution data assumptions. Fractional power is estimated at 2, 3, & 4σ standard deviation population levels above a reference normal  $N(0,1)$  background population. These results are compared with Unified Guidance reference criteria at 3 and 4 σ.

The initial version of this program (September 2004, Version 1) utilized regression equations of Monte Carlo simulation power data. While reasonable approximations, it was discovered that the fractional power estimates did not closely match the available draft Unified Guidance transition cutoff levels (between categorical power ratings for the maximum and 2<sup>nd</sup> maximum).

The current program version follows the approach in the September 2004 draft Unified Guidance-- ".... a simple modification to the numerical algorithms presented in Davis and McNichols (1987) <sup>4</sup> was used to compute power directly." The calculator uses a simple modification of their basic parametric equation, converted to fixed maximal values approximating a normal distribution for non-parametric 1:*m* tests. Davis and McNichols derived the equation below for parametric normal prediction limits containing the following inputs: a background sample size *n*, *r* successive comparisons, *p*-of-*m* type tests and some *K*-factor based on the background statistics of the form  $\bar{x} + K \cdot s$ :

$$\Pr(Y^* < K \cdot s) = \int_0^1 T_{n-1, \sqrt{n}[\Phi^{-1}(v) + K/\sigma]} \left( \sqrt{n} K \right) \cdot r I^{r-1}(v, p, m+1-p) \frac{v^{p-1} (1-v)^{m-p}}{B(p, m+1-p)} \cdot \partial v$$

. While this equation appears inordinately complex, all but the non-central *T*-function can be easily handled by routine statistical software or even simpler computations. The essence of the equation is to integrate from 0 to 1 for the non-central *T* with appropriate inputs, the cumulative Beta distribution  $I(\bullet, a, b)$  multiplied by *r* and taken to the *r*-1 exponent, and the Beta pdf (the last ratio terms in the equation). For 1:*m* tests, the two Beta functions reduce to simple algebraic computations. The standard normal distribution,  $\Phi$ , was also readily estimated with

---

<sup>4</sup> Davis, Charles B. & McNichols, Roger J., "One-Sided Intervals for at Least *p* of *m* Observations From a Normal Population on Each of *r* Future Occasions", Technometrics, August 1987, pp. 359-370

common programming techniques. As a further simplification,  $r = 1$  for all cases, following the 2004 and current 2009 guidance approach of evaluating power based on a single future test.

In the present program, a fairly large number of  $v$  iterations for the probability range (0,1) are first generated in an initial column as  $(i - .5) / 1000$  for  $i$  from 1 to 1000. The inverse normal distribution function [ $\Phi^{-1}(v)$ ] then assigns  $Z$ -values to these probabilities. The Beta pdfs and cdfs are then calculated for the same range and stored based on the appropriate input parameters  $v$ ,  $p$ , and  $m$ . Calculated  $rI^{r-1}$  results sum to 1.0.

The non-central  $T$  poses the most difficult computation problem, since it must be assessed at various  $\sqrt{n} \cdot [Z(v) + \Delta/\sigma]$  values (with  $Z(v)$  varying from  $v = 0$  to 1) against a factor comparable to  $\sqrt{n} \cdot K$ . To keep computer running time low and easily understood, the Hald approximating equation was used for the non-central  $T$  in this algorithm. Hald<sup>5</sup> found that for reasonably low sample size, the non-central  $T$  probability can be approximated by the standard normal distribution utilizing a central- $t$  distribution function as follows:

$$\pi(1 - \beta) = \Phi \left| \frac{\lambda - t_{1-\alpha, N-1} \cdot \left(1 - \frac{1}{4(N-1)}\right)}{\sqrt{1 + \frac{t_{1-\alpha, N-1}^2}{2 \cdot (N-1)}}} \right|$$

where  $\pi(1 - \beta)$  is the fractional power probability,  $N - 1$  is the degrees of freedom,  $t$  is a central  $t$  at  $N - 1$  degrees of freedom and  $1 - \alpha$  confidence,  $\Phi[*]$  is the standard normal distribution probability value, and  $\lambda$  is a non-central- $T$  parameter specific to the type of comparisons made. The approximating equation is quite accurate for assessing most power situations, but least so at  $\Delta = 0$  or very low positive  $\Delta$ . The Hald normality assumption works adequately in most situations for sample sizes of  $n \geq 6-8$ .

Since the Optimal Rank Values Calculator program generates fixed maximal values from a given background sample size, these must first be converted to a probability, which can then be converted to a  $Z$ -value, assuming that both the background and power distributions are normal. In order to identify the appropriate comparison value for the non-central  $T$  statistic, the following approach was used. Given some maximal rank of  $x_{\max}$  as the  $y$ -th largest value from a background sample of size  $n$  used as the prediction limit, the rank is first converted to a probability of occurrence  $p$  given some fractional (0 - 1) value  $f$  as follows:

$$p = \frac{(y - [1 - f] / 2)}{n + f}$$

---

<sup>5</sup>

Hald, Anders, Statistical Theory with Engineering Applications, John Wiley & Sons, Inc., 1952, Chapter 15, pp. 391-393

This probability is then converted to a Z- value using the inverse normal function. After some experimentation, it was found that the best approximation was  $f=0$ , or  $p = (y - .5)/n$ . The corresponding non-parametric Hald  $t$ -comparison value for  $\sqrt{n}K$  in the Davis & McNichols Equation 1 is  $\sqrt{n} \cdot Z(p)$ . The lambda ( $\lambda$ ) non-central t parameters are  $\sqrt{n} \cdot [Z(v) + \Delta/\sigma]$ , with  $v$  varying from 0 to 1.

The resulting 1000 probability values for the non-central  $T$  are then multiplied against the respective Beta pdf and  $rI^{r-1}$  (cdf) values for each probability from 0 to 1. The numerical fractional probability is then the simple integration achieved by summing these resultant products times the single (1/1000) delta integration value. The same algorithm is followed varying  $\sigma$  power levels from 2 to 4 for each 1: $m$  test.

While specific fractional power values for the non-parametric tests were not available in the guidance for comparison, the effectiveness of these two calculator power approaches (Version 1 regression information and the more recent approximation of the Davis & McNichols equation) was assessed in two ways. The first assessment compares transition cutoff levels in the September 2004 guidance appendix Tables 13-19 through 13-24. The second approach is to compare the current Optimal Rank calculator outputs with the earlier Monte Carlo data.

In the 2004 and more recent 2009 Unified Guidance, power is assessed at  $3\sigma$  and  $4\sigma$ , utilizing a one-way upper 99% normal 1:1 prediction interval test based on a sample size  $n=10$  as the reference power curve. If fractional test power at  $4\sigma$  lies below the reference value (.815), power is rated as "Low". If above, the rating is "Acceptable". If the  $3\sigma$  test power lies above .541 as well, then the rating is "Good". Guidance tables identify these three categories respectively as plain text, italicized, or bolded; no actual fractional power is provided. To maintain consistency in comparisons, the EPA reference test criteria were calculated as .539 at  $3\sigma$  and .814 for  $4\sigma$  using a parametric normal power approximation algorithm similar to that in the Optimal Rank Values Calculator program. The latter reference values have been used in the two calculator versions to assess categorical power ratings. The table below provides the background sample size transition cutoff level comparisons for 1:2 through 1:4 non-parametric tests:

| Comparison of Background Sample Size Power Rating Transition Cutoff Levels |                            |               |                |                |                         |          |          |
|--|----------------------------|---------------|----------------|----------------|-------------------------|----------|----------|
| Test Type  | Sigma Level(s)             | Maximum Value |                |                | 2 <sup>nd</sup> Maximum |          |          |
|  |                            | UG            | ORC-2004       | ORC-2006       | UG                      | ORC-2004 | ORC-2006 |
| 1:2  | 3&4 $\sigma$<br>4 $\sigma$ | 61-70         | <b>77-78</b>   | <b>57-58</b>   | 161-180                 | 170-171  | 172-173  |
|  |                            | 121-140       | <b>160-161</b> | <b>142-143</b> | > 200                   | >200     | >200     |
| 1:3  | 3&4 $\sigma$<br>4 $\sigma$ | 31-35         | <b>39-40</b>   | 29-30          | 81-90                   | 88-89    | 86-87    |
|  |                            | 71-80         | <b>92-93</b>   | 74-75          | >200                    | >200     | >200     |
| 1:4  | 3&4 $\sigma$<br>4 $\sigma$ | 21-25         | 25-26          | 19-20          | 51-60                   | 57-58    | 56-57    |
|  |                            | 41-50         | <b>62-63</b>   | 49-50          | 141-160                 | 143-144  | 158-159  |

"ORC-2004" refers to the initial Optimal Rank Calculator version using power regression data and the second "ORC-2006" the more recent program using the Hald non-central  $t$  approximation. The UG transition levels span somewhat larger ranges, since the guidance tables provided limited background sample size inputs. With the Optimal Rank Calculator programs, it was possible to identify the transition ranges to a difference of one sample size.

Bolded values indicate where the calculator programs differed from the guidance ranges by more than a single sample size. While both calculator programs worked well for the 2<sup>nd</sup> maximum, the results differed for the maximum value, especially for the 1:2 test. The recent ORC-2006 algorithm more closely approximates the guidance levels, which were generated with a formal non-central  $T$  program. None of the latter transition ranges differ more than 1-3 sample size values from the Unified Guidance ranges.

The second comparison is to the original Monte Carlo data. The earlier calculator version regression results are less accurate and not provided here. Monte Carlo simulations were used to generate power results at 2, 3 and 4 $\sigma$  levels for 1:1 through 1:4 prediction limit tests based on fixed background ranks. Background samples were generated at various  $n$  from 6 to 199. Maximal (inverse) ranks ranged from 1 to 4 for smaller background sizes to 1 to 12 for larger  $n$ . A total of 1392 fractional power values were generated under these test conditions. The Optimal Rank calculator was then used to generate the corresponding fractional power levels at the same input conditions. The table below provides basic statistics for differences between the calculator and Monte Carlo values. Both aggregate statistics and those for the 1<sup>st</sup> maximum compared to the remaining ranks is provided:

| <b>Comparison of OptRankCalc and Monte Carlo Power Data<br/>(Difference of OptRank - MC)</b> |                   |                               |                      |
|--|-------------------|-------------------------------|----------------------|
| <b>Statistic</b>   | <b>Categories</b> |                               |                      |
|  | <b>Aggregate</b>  | <b>1<sup>st</sup> Maximum</b> | <b>Remaining Max</b> |
| <b>N</b>   | 1392              | 192                           | 1200                 |
| <b>Mean, <math>\bar{x}</math></b>  | -.00265           | -.0187                        | -.00008              |
| <b>Median</b>  | 0                 | -.018                         | .001                 |
| <b>Std. Dev</b>  | .0122             | .0218                         | .00698               |
| <b>Minimum</b>   | -.067             | -.067                         | -.043                |
| <b>Maximum</b>   | .041              | .018                          | .041                 |
| <b>% <math>\leq</math> .05</b>   | 98.7              | 90.6                          | 100.0                |
| <b>% <math>\leq</math> .025</b>  | 93.6              | 57.3                          | 99.4                 |
| <b>% <math>\leq</math> .01</b>   | 82.0              | --                            | 89.4                 |

With the exception of the absolute maximum rank, the algorithm used in the Optimal Rank Calculator closely tracks the Monte Carlo power data for rank values as low as the 12th

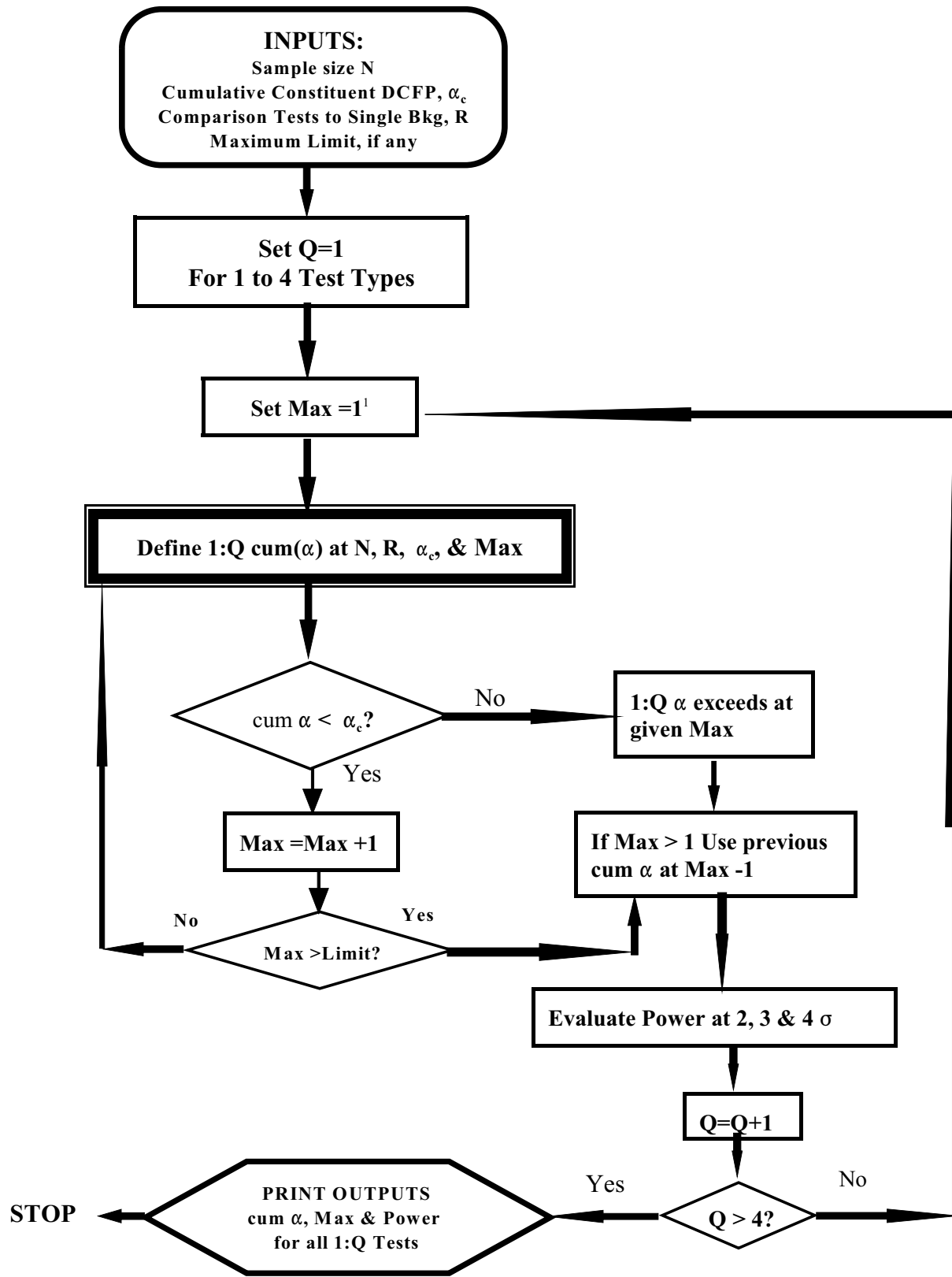
highest. Almost 90% of the algorithm values lie within  $\pm .01$  power units. For most design conditions, the calculator power results appears quite adequate. As noted in the first comparison with transition cutoff levels, the 2006 calculator power algorithm is still somewhat inexact at the maximum rank. A further evaluation of the 192 data points for the maximum value showed that at  $4\sigma$ , the algorithm estimate of differences was quite close to Monte Carlo results ( $\bar{x} = .0033$ ;  $s = .0094$ ). However, the 1<sup>st</sup> rank estimates were low at both  $2$  and  $3\sigma$  ( $\bar{x} = -.03$ ;  $s = .017$ ). This may somewhat affect interpretations using the 1:2 test at the maximum rank (since the 1:1 test isn't usually applicable, and 1:3 or 1:4 tests generally are optimal at lower ranks). Overall, these tabular results confirm the accuracy of calculator power estimates for all other background value ranks.

Since EPA power comparisons are only relative to a standard normal prediction interval in specified power ranges, the degree of accuracy at extremely high or low power isn't particularly critical. Background or downgradient well distributions other than the normal could change the power outcomes, but the present effort and EPA guidance were primarily intended to establish an understandable, systematic and repeatable power evaluation. Using the calculator avoids the need to run additional Monte Carlo simulations for these types of thoroughly-evaluated 1:m non-parametric prediction interval tests.

Using the calculator information, it is possible to generate specific effect size power estimates based on the fractional power outputs at  $\Delta = 0, 2, 3$  &  $4\sigma$  for each 1:m test. For a specific effect size:  $\Delta_{SE} = (G - \bar{x})/s$  at some mean level  $G$  above a sample background with statistics  $\bar{x}$  and  $s$ , either the power values can be graphed against  $\Delta$  and the power estimated by eye, or linearly transformed and regressed (such as a logarithmic transformation:  $\log\{[1/\text{fractional power}] - 1\}$  versus  $\Delta$ ). The effect size power can then be estimated from the regression equation at  $\Delta_{SE}$ . Results will only be approximate, and the caveats discussed above for the maximum rank need to be kept in mind.

Figures

FIGURE 1. FLOW DIAGRAM FOR ‘Optimal Rank Values Calculator’ PROGRAM

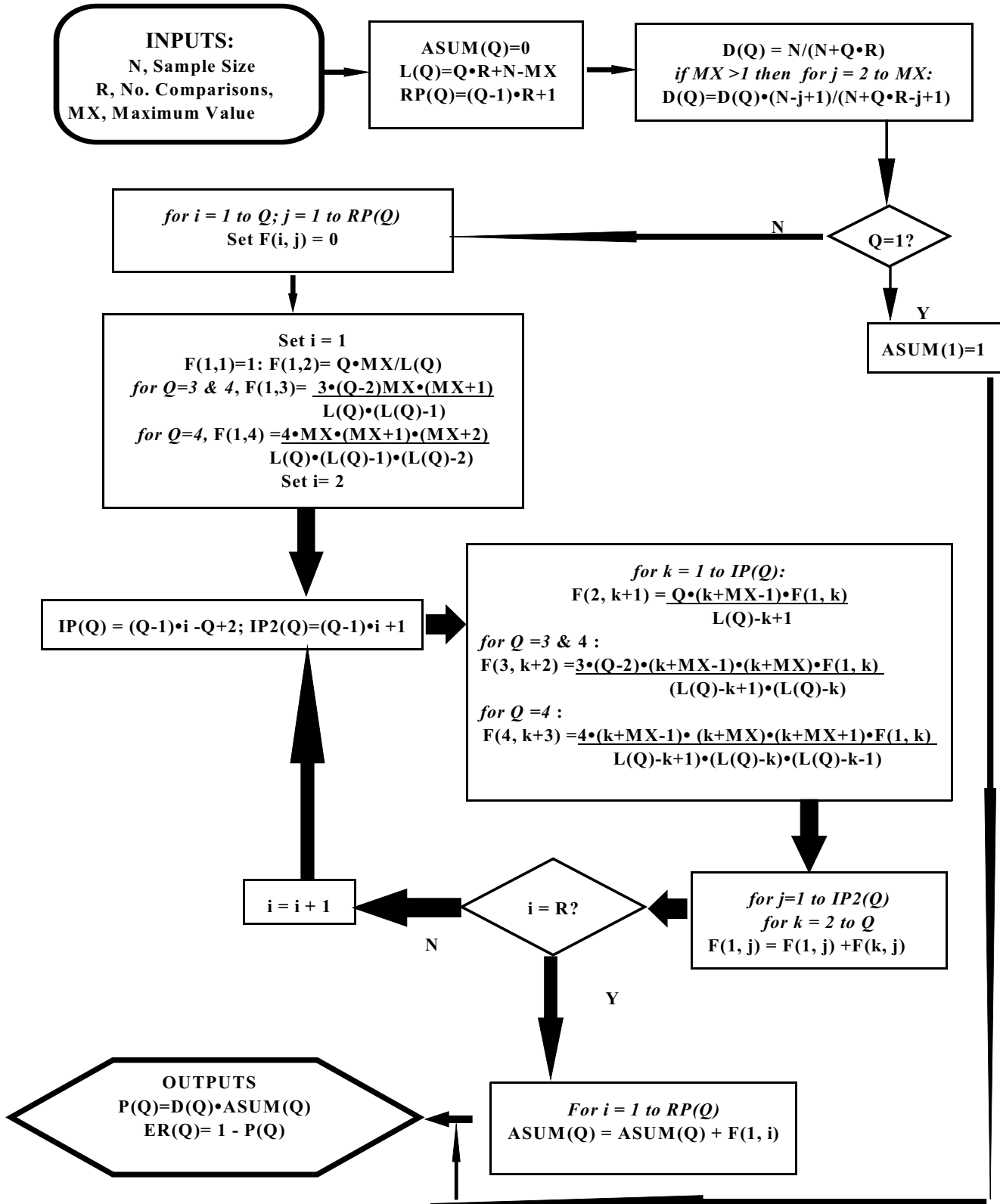


<sup>1</sup> For increased efficiency, initial Max(Q) set at previous Max(Q-1)



FIGURE 2

FLOW DIAGRAM for 1:Q NPM CUMULATIVE FALSE POSITIVE  $\alpha$  (Q=1 to 4)



**FIGURE 3**  
**METHOD OF RECURSIVE CALCULATION FOR NON-PARAMETRIC 1 of  $m$  TESTS**

**Test Type: 1:1**

Inputs: N=5 (sample size) R=3 (# Comparison Tests) M=2 (maximum)

Precalculations: D1(1)=N/(N+R) = 5/8

For j = 2 to M D1(j) = D1(j-1)\*(N-j+1)/(N+R-j+1) = 5/8\*4/7 = 5/14

Calculations: Cum P1(M)=D1(M) = 5/14; **Cum Q1= 1- P1(M) = 9/14 = .6429**

**Test Type: 1:2**

Inputs: N=5 (sample size) R=3 (# Comparison Tests) M=2 (maximum)

Precalculations: D2(1)=N/(N+2R) = 5/11

For j = 2 to M D2(j) = D2(j-1)\*(N-j+1)/(N+2R-j+1) = 5/11\*4/10 = 2/11

L=N+2R-M = 9; RP = 2+ (R-1) = 4

**Initialization Row for R=1 for K = 1 to RP**

| K =               | 1 | 2              | 3 | 4= RP |
|-------------------|---|----------------|---|-------|
| F(1,K)<br>Formula | 1 | $\frac{2M}{L}$ | 0 | 0     |
| F(1,K)            | 1 | $\frac{4}{9}$  | 0 | 0     |

**Multiplier F(2, K+1) = F(1,K)• [2•(K+M-1)/(L-K+1)]**

|                      |   |               |               |               |
|----------------------|---|---------------|---------------|---------------|
| F(1,k)<br>Multiplier | * | $\frac{4}{9}$ | $\frac{6}{8}$ | $\frac{8}{7}$ |
|----------------------|---|---------------|---------------|---------------|

**Calculations for R = 2**

|                |   |                                     |   |   |
|----------------|---|-------------------------------------|---|---|
| F(1,K)         | 1 | $\frac{4}{9}$                       | 0   | 0 |
| F(2,K)         | 0 | $1 \cdot \frac{4}{9} = \frac{4}{9}$ | $\frac{6}{8} \cdot \frac{4}{9} = \frac{1}{3}$ | 0 |
| $\sum (1 + 2)$ | 1 | $\frac{8}{9}$                       | $\frac{1}{3}$                                 | 0 |

**Calculations for R = 3**

|                |   |                                     |   |  |
|----------------|---|-------------------------------------|---|--|
| F(1,K)         | 1 | $\frac{8}{9}$                       | $\frac{1}{3}$                                 | 0  |
| F(2,K)         | 0 | $1 \cdot \frac{4}{9} = \frac{4}{9}$ | $\frac{6}{8} \cdot \frac{8}{9} = \frac{2}{3}$ | $\frac{8}{7} \cdot \frac{1}{3} = \frac{8}{21}$ |
| $\sum (1 + 2)$ | 1 | $\frac{12}{9}$                      | 1   | $\frac{8}{21}$                                 |

Calculations:  $\sum \sum (1+2) \text{ Row} = (1 + 12/9 + 1 + 8/21) = 702/189$

Cum P2 =  $\sum \sum (1+2) \text{ Row} \times D2 = (702/189) \times (2/11) = .6753$ ; **Cum Q2= 1-P2 = .3247**

**FIGURE 3 (cont'd)**

**Test Type: 1:3**

Inputs: N=5 (sample size) R=3 (# Comparison Tests) M=2 (maximum)  
Precalculations: D3(1)=N/(N+3R) = 5/14  
For j = 2 to M D3(j) = D3(j+1)\*(N-j+1)/(N+3R-j+1) = 5/14\*4/13 = 10/91  
L=N+3R-M = 12; RP = 3+ 2(R-1) = 7

**Initialization Row for R=1 for K = 1 to RP**

| K =                        | 1 | 2              | 3                        | 4 | 5 | 6 | 7=RP |
|----------------------------|---|----------------|--------------------------|---|---|---|------|
| <b>F(1,K)<br/>Formulae</b> | 1 | $\frac{3M}{L}$ | $\frac{3M(M+1)}{L(L-1)}$ | 0 | 0 | 0 | 0    |
| <b>F(1,K)</b>              | 1 | $\frac{1}{2}$  | $\frac{3}{22}$           | 0 | 0 | 0 | 0    |

**F(2, K+1) = F(1, K)•[3•(K+M-1)/(L-K+1)]**

|  |   |               |                |               |               |               |                |
|--|---|---------------|----------------|---------------|---------------|---------------|----------------|
| <b>1<sup>st</sup> F(1, K)<br/>Multiplier</b> | * | $\frac{1}{2}$ | $\frac{9}{11}$ | $\frac{6}{5}$ | $\frac{5}{3}$ | $\frac{9}{4}$ | $\frac{21}{7}$ |
|--|---|---------------|----------------|---------------|---------------|---------------|----------------|

**F(3, K+2) = F(1, K)•[3•(K+M-1)•(K+M)]/[(L-K+1)•(L-K)]**

|  |   |   |                |                 |               |               |               |
|--|---|---|----------------|-----------------|---------------|---------------|---------------|
| <b>2<sup>nd</sup> F(1, K)<br/>Multiplier</b> | * | * | $\frac{3}{22}$ | $\frac{18}{55}$ | $\frac{2}{3}$ | $\frac{5}{4}$ | $\frac{9}{4}$ |
|--|---|---|----------------|-----------------|---------------|---------------|---------------|

**Calculations for R = 2**

|                     |   |               |                 |                 |                |   |   |
|---------------------|---|---------------|-----------------|-----------------|----------------|---|---|
| <b>F(1,K)</b>       | 1 | $\frac{1}{2}$ | $\frac{3}{22}$  | 0               | 0              | 0 | 0 |
| <b>F(2,K)</b>       | 0 | $\frac{1}{2}$ | $\frac{9}{22}$  | $\frac{9}{55}$  | 0              | 0 | 0 |
| <b>F(3,K)</b>       | 0 | 0             | $\frac{3}{22}$  | $\frac{9}{55}$  | $\frac{1}{11}$ | 0 | 0 |
| <b>∑ (1 + 2 +3)</b> | 1 | 1             | $\frac{15}{22}$ | $\frac{18}{55}$ | $\frac{1}{11}$ | 0 | 0 |

**Calculations for R = 3**

|                     |   |               |                 |                 |                 |                 |                |
|---------------------|---|---------------|-----------------|-----------------|-----------------|-----------------|----------------|
| <b>F(1,K)</b>       | 1 | 1             | $\frac{15}{22}$ | $\frac{18}{55}$ | $\frac{1}{11}$  | 0               | 0              |
| <b>F(2,K)</b>       | 0 | $\frac{1}{2}$ | $\frac{9}{11}$  | $\frac{9}{11}$  | $\frac{6}{11}$  | $\frac{9}{44}$  | 0              |
| <b>F(3,K)</b>       | 0 | 0             | $\frac{3}{22}$  | $\frac{18}{55}$ | $\frac{5}{11}$  | $\frac{9}{22}$  | $\frac{9}{44}$ |
| <b>∑ (1 + 2 +3)</b> | 1 | $\frac{3}{2}$ | $\frac{18}{11}$ | $\frac{81}{55}$ | $\frac{12}{11}$ | $\frac{27}{44}$ | $\frac{9}{44}$ |

Calculations:  $\sum \sum (1+2+3) \text{ Row} = (1 + \frac{3}{2} + \frac{18}{11} + \frac{81}{55} + \frac{12}{11} + \frac{27}{44} + \frac{9}{44}) = 7.5182$   
Cum P3 =  $\sum \sum (1+2+3) \text{ Row} \times D3 = (7.5182) \times (\frac{10}{91}) = .8262$ ; **Cum Q3= 1-P3 = .1738**

-----

**FIGURE 3(cont'd)**

**Test Type: 1:4**

Inputs: N=5 (sample size) R=3 (# Comparison Tests) M=2 (maximum)  
Precalculations: D4(1)=N/(N+4R) = 5/17  
For j = 2 to M D4(j) = D4(j-1)\*(N-j+1)/(N+4R-j+1) = 5/17\*4/16 = 5/68  
L=N+4R-M = 15; RP = 4+ 3(R-1) = 10

**Initialization Row for R=1 for K = 1 to RP**

| K =                        | 1 | 2              | 3                        | 4                                  | 5 | 6 | 7 | 8 | 9 | 10=RP |
|----------------------------|---|----------------|--------------------------|------------------------------------|---|---|---|---|---|-------|
| <b>F(1,K)<br/>Formulae</b> | 1 | $\frac{4M}{L}$ | $\frac{6M(M+1)}{L(L-1)}$ | $\frac{4M(M+1)(M+2)}{L(L-1)(L-2)}$ | 0 | 0 | 0 | 0 | 0 | 0     |
| <b>F(1,K)</b>              | 1 | 8/15           | 6/35                     | 16/455                             | 0 | 0 | 0 | 0 | 0 | 0     |

**F(2, K+1) = F(1, K) • [4•(K+M-1)/(L-K+1)]**

|                                  |   |      |     |       |     |       |      |      |     |      |
|----------------------------------|---|------|-----|-------|-----|-------|------|------|-----|------|
| <b>1<sup>st</sup> Multiplier</b> | * | 8/15 | 6/7 | 16/13 | 5/3 | 24/11 | 14/5 | 32/9 | 9/2 | 40/7 |
|----------------------------------|---|------|-----|-------|-----|-------|------|------|-----|------|

**F(3, K+2) = F(1, K) • [6•(K+M-1)•(K+M)]/[L(L-K+1)•(L-K)]**

|                                  |   |   |      |       |       |       |        |       |   |        |
|----------------------------------|---|---|------|-------|-------|-------|--------|-------|---|--------|
| <b>2<sup>nd</sup> Multiplier</b> | * | * | 6/35 | 36/91 | 10/13 | 15/11 | 126/55 | 56/15 | 6 | 135/14 |
|----------------------------------|---|---|------|-------|-------|-------|--------|-------|---|--------|

**F(4, K+3) = F(1, K) • [4•(K+M-1)•(K+M)•(K+M+1)]/[L(L-K+1)•(L-K)•(L-K-1)]**

|                                  |   |   |   |        |       |        |      |         |      |      |
|----------------------------------|---|---|---|--------|-------|--------|------|---------|------|------|
| <b>3<sup>rd</sup> Multiplier</b> | * | * | * | 16/455 | 10/91 | 40/143 | 7/11 | 224/165 | 14/5 | 40/7 |
|----------------------------------|---|---|---|--------|-------|--------|------|---------|------|------|

**Calculations for R = 2**

| K =                 | 1 | 2               | 3               | 4                 | 5                | 6                 | 7                  | 8 | 9 | 10=RP |
|---------------------|---|-----------------|-----------------|-------------------|------------------|-------------------|--------------------|---|---|-------|
| <b>F(1,K)</b>       | 1 | 8/15            | 6/35            | 16/455            | 0                | 0                 | 0                  | 0 | 0 | 0     |
| <b>F(2,K)</b>       | 0 | 8/15            | 16/35           | 96/455            | 16/273           | 0                 | 0                  | 0 | 0 | 0     |
| <b>F(3,K)</b>       | 0 | 0               | 6/35            | 96/455            | 36/273           | 48/1001           | 0                  | 0 | 0 | 0     |
| <b>F(4,K)</b>       | 0 | 0               | 0               | 16/455            | 16/273           | 48/1001           | $\frac{112}{5005}$ | 0 | 0 | 0     |
| <b>∑<br/>(1..4)</b> | 1 | $\frac{16}{15}$ | $\frac{28}{35}$ | $\frac{224}{455}$ | $\frac{68}{273}$ | $\frac{96}{1001}$ | $\frac{112}{5005}$ | 0 | 0 | 0     |

**Calculations for R = 3**

|                     |   |                 |                 |                   |                   |                     |                     |                       |                       |                        |
|---------------------|---|-----------------|-----------------|-------------------|-------------------|---------------------|---------------------|-----------------------|-----------------------|------------------------|
| <b>F(1,K)</b>       | 1 | $\frac{16}{15}$ | $\frac{28}{35}$ | $\frac{224}{455}$ | $\frac{68}{273}$  | $\frac{96}{1001}$   | $\frac{112}{5005}$  | 0                     | 0                     | 0                      |
| <b>F(2,K)</b>       | 0 | $\frac{8}{15}$  | $\frac{32}{35}$ | $\frac{448}{455}$ | $\frac{224}{273}$ | $\frac{544}{1001}$  | $\frac{1344}{5005}$ | $\frac{3584}{45045}$  | 0                     | 0                      |
| <b>F(3,K)</b>       | 0 | 0               | $\frac{6}{35}$  | $\frac{192}{455}$ | $\frac{168}{273}$ | $\frac{672}{1001}$  | $\frac{2856}{5005}$ | $\frac{16128}{45045}$ | $\frac{6048}{45045}$  | 0                      |
| <b>F(4,K)</b>       | 0 | 0               | 0               | $\frac{16}{455}$  | $\frac{32}{273}$  | $\frac{224}{1001}$  | $\frac{1568}{5005}$ | $\frac{15232}{45045}$ | $\frac{12096}{45045}$ | $\frac{40320}{315315}$ |
| <b>∑<br/>(1..4)</b> | 1 | $\frac{24}{15}$ | $\frac{66}{35}$ | $\frac{880}{455}$ | $\frac{492}{273}$ | $\frac{1536}{1001}$ | $\frac{5880}{5005}$ | $\frac{34944}{45045}$ | $\frac{18144}{45045}$ | $\frac{40320}{315315}$ |

Calculations:  $\sum \sum(1..4) = 12.237$ ; Cum P4 = 12.237\*(5/68) = .8998 **Cum Q4 = 1- .8998 = .1002**